NTNU
Norwegian University of
Science and Technology

Computing Metrics on Riemannian Shape Manifolds

Geometric shape analysis made practical

Eivind Fonn

Master of Science in Physics and Mathematics
Submission date: June 2009
Supervisor: Elena Celledoni, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

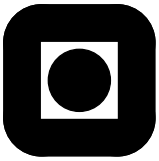
Problem Description

We will study an approach to the analysis of two-dimensional shapes based on discretization of infinitely-dimensional Riemannian manifolds, so called "loop spaces". We will develop algorithms for solving several central problems, such as deciding "closeness" between shapes, and continuously deforming one shape into another. The algorithms will be tested on sample problems, and their running time and efficiency will be analyzed. We will also look into using alternative representations for shapes and determine their relative merits. Some real-world applications will be considered.

Assignment given: 15. January 2009
Supervisor: Elena Celledoni, MATH

NTNU
Norwegian University of Science and Technology

Faculty of Information Technology, Mathematics and Electrical Engineering



MASTER'S THESIS

FOR

STUD.TECHN. Eivind Fonn

FACULTY OF INFORMATION TECHNOLOGY, MATHEMATICS AND ELECTRICAL ENGINEERING

NTNU

Date due: June 11, 2009

Discipline: Numerical analysis

"Computing Metrics on Riemannian Shape Manifolds"

Purpose of the work: We will study an approach to the analysis of two-dimensional shapes based on discretization of infinite-dimensional Riemannian manifolds, so-called "loop spaces". We will develop algorithms for solving several central problems, such as deciding "closeness" between shapes, and continuously deforming one shape into another. The algorithms will be tested on sample problems, and their running time and efficiency will be analyzed. We will also look into using alternative representations for shapes and determine their relative merits. Some real-world applications will be considered.

This diploma thesis is to be carried out at the Department of Mathematical Sciences under guidance of Associate Professor Elena Celledoni.

Trondheim, January 15, 2009.

Trond Digernes
Instituttleder
Dept. of Mathematical Sciences

Elena Celledoni
Associate Professor
Dept. of Mathematical Sciences

Abstract

Shape analysis and recognition is a field ripe with creative solutions and innovative algorithms. We give a quick introduction to several different approaches, before basing our work on a representation introduced by Klassen et. al. [11], considering shapes as equivalence classes of closed curves in $\mathbb{R}^2$ under reparametrization, and invariant under translation, rotation and scaling. We extend this to a definition for nonclosed curves, and prove a number of results, mostly concerning under which conditions on γ the set of shapes become manifolds. We then motivate the study of geodesics on these manifolds as a means to compute a shape metric, and present two methods for computing such geodesics: the shooting method from [11] and the "direct" method, new to this paper. Some numerical experiments are performed, which indicate that the direct method performs better for realistically chosen parameters, albeit not asymptotically.

Contents

1 Notation

S^1	The 1-dimensional sphere, the unit circle in $\mathbb{R}^2$ or, equivalently, $\mathbb{R}/2\pi\mathbb{Z}$. S^1 is an abelian Lie group under addition. (6)
τf	For $f \in L^2(0, 2\pi)$, τf is the mirror function $\tau f(t) = f(2\pi - t)$. (6)
$L^2(0, 2\pi)$	The Hilbert space of square integrable functions on $(0, 2\pi)$, equipped with the typical inner product $\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)\,\mathrm{d}t$. (7)
$\mathbb{Z}_2$	The (only) group of order 2. We use the notation $\mathbb{Z}_2 = \{\tau, \varepsilon\}$, where ε is the identity and $\tau^2 = \varepsilon$, to correspond with the notation τf above. (7)
L_g	For a left group action from G on M, L_g is the left-multiplication map by g: $L_g(x) = g \cdot x$. (7)
Df, $Df\|_x$	For a smooth map f, Df is its pushforward, and $Df\|_x$ is the restriction of the pushforward to the tangent space at x. (7)
$d_X(\cdot, \cdot)$	Metric on the space X, which in this context is usually a Riemannian manifold. (8)
$[x]$	Equivalence class containing x. (8)
id_X	Identity map on a set X. (8)
$S_p \subset L^2(0, 2\pi)$	The space of 2π-periodic smooth functions on $\mathbb{R}$, equipped with the induced inner product from $L^2(0, 2\pi)$. (8)
S_{sp}	The affine space $\mathrm{id}_{\mathbb{R}} + S_p$ of semiperiodic functions. (8)
$c \cdot f$	For $f \in S_{sp}$ and $c \in \mathbb{R}$, the function $c \cdot f(t) = c \odot f(t) - c = f(c+t) - c$. For $f \in \mathbb{R}^N$ and $c \in \mathbb{Z}$, extend f semiperiodically to $\mathbb{Z}$, and $\mathrm{let}(c \cdot f)^i = f^{i+c} - hc$, where Nh is the elevation of the semiperiodic extension. See footnote on page 16. (8)
$c \odot f$	For a function f and $c \in \mathbb{R}$, the function $c \odot f(t) = f(t + c)$. (10)
$T_x\mathcal{M}$	The tangent space of the manifold $\mathcal{M}$ at x. (10)
$N_x\mathcal{M}$	The normal space of a Riemannian submanifold $\mathcal{M} \subset \mathcal{N}$ at x, where $\mathcal{N}$ is usually understood. (10)
σ_x	A retraction at x. See definition 15. (14)
P_x	With $x \in \mathcal{M} \subset \mathcal{N}$, where $\mathcal{M}, \mathcal{N}$ are Riemannian manifolds, P_x is the orthogonal projection $T_x\mathcal{N} \to T_x\mathcal{M}$. (14)
$\mathbb{1}_N$	The column N-vector $\mathbb{1}_N = (1, \ldots, 1)^T$. (15)
$\exp_x$	For a Riemannian manifold $\mathcal{M}$, $\exp_x$ is the exponential mapping $U \subset T_x\mathcal{M} \to \mathcal{M}$ at x. (15)
$\mathbb{Z}_N$	The additive group of integers modulo N, $\mathbb{Z}_N \cong \mathbb{Z}/n\mathbb{Z}$. (16)
$A \doteq B$	A is, by definition, equal to B. (17)
I_M	The $M \times M$ identity matrix. (19)

2 Introduction

Shape analysis is a relatively new field on the borderline between mathematics and computer science. At the core we find the problem of *shape recognition*: we strive to automatize a process which the human brain can perform with almost annoying ease: deciding when two shapes are the same, or if they are not, how equal they are.[1] Not surprisingly, perhaps, we find ourselves in a familiar situation: a process which our brains can do so easily, our computers find very challenging.

There is ample motivation for solving this problem in a satisfactory manner. Aside from the obvious motivation in automatizing such a natural process, such a program will have many applications in the medical sciences. Image processing plays a central role in diagnosis, and due to the demand for high accuracy, most of this is done manually, at great expense in time and money. See for example [2]. One of the central subfields is Computational Anatomy, or CA, discussed in [7]. Other curious applications can be mentioned, such as paleontology [17] and custom fitting of footwear [20].

One particularly interesting question is how to *represent* a shape. Numerous representations have been proposed, and doubtless the most fitting one will depend on context. A good shape representation should answer to some of the following conditions:

1. The representation should be easy to compute from the "shape" (whatever that is) and vice versa.

2. The representation should be sufficiently transparent to allow ease of analysis. A representation which is too complicated will not make our work easier.

3. Certain actions on shapes, such as translation, rotation and scaling, to name a few, should be easily modeled in terms of the representation. This will allow us to form quotient sets of shapes. Depending on the context, we might want to consider shapes as equivalent under one or more of these actions.

4. The representation should not be inherently numerically challenging. In particular, one should avoid having to work with high-dimensional diffeomorphisms.

We present a short overview of some interesting representations.

2.1 Characteristic functions, distance functions

Let $E \subset \mathbb{R}^2$ be a set, either a set we usually identify as a "shape", or any other set. We then have the characterestic function of E,

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{otherwise,} \end{cases}$$

and the distance function,

$$d_E(x) = \inf_{y \in E} \|x - y\|.$$

[1] At least relatively: is shape A closer to B than C is to D?

The characteristic function is always a faithful representation of E, but it is in general discontinuous. The distance function *is* continuous, and if E is closed, we have $E = d_E^{-1}(0)$. Basing our representation on d_E is tempting, as there is no need to consider reparametrizations, and the variety of possible sets E is nothing short of huge. On the other hand, it seems to us that quotienting out actions like rotation, scaling and translation is not straightforward, computationally speaking.

This approach has been pursued with great rigor in [6].

2.2 Harmonic embeddings

Here, we only consider closed two-dimensional curves in an annulus $\Omega = \{x \in \mathbb{R}^2 \mid r < \|x\| < R\}$, which are zero sets of some harmonic function u on Ω, i.e. $\Delta u = 0$ and $E = u^{-1}(0)$ is the given shape. The idea is that due to the uniqueness of solutions of Laplace's equation, the function u is given uniquely by its values on the inner and outer boundaries of Ω. If the boundary values are of opposite sign, E is always simply connected.

The shape space, built from the periodic functions on the boundaries, is now linear in a very natural way. See [9] for details.

2.3 Conformal mappings

With the usual identification between $\mathbb{R}^2$ and $\mathbb{C}$, we consider shapes as smooth, closed and simple curves in $\mathbb{C}$. By the Riemann mapping theorem, given any such curve γ, there is a diffeomorphism of $\mathbb{C}$ taking the unit disk to the interior of γ. This diffeomorphism is unique up to precomposition of a subset of Möbius maps: those taking the unit disk to itself (of the form $z \mapsto (az + b)/(\bar{b}z + \bar{a})$). Quotienting this action out yields a unique representation in terms of equivalence classes of diffeomorphisms of $\mathbb{C}$. It is notable, however, that such diffeomorphisms are very challenging, numerically speaking. See [18].

2.4 Paths of deformations

This approach considers not shapes, but *images*, a related concept. Given two (grayscale, say) images $I_1, I_2 : \mathbb{R}^2 \to \mathbb{R}$, we seek a path of deformations $\phi_s : \mathbb{R}^2 \to \mathbb{R}^2$, $0 \leq s \leq 1$, so that ϕ_0 is the identity, and $I_2 = I_1 \circ \phi_1$. The set of diffeomorphisms of $\mathbb{R}^2$ is an infinite dimensional Lie group, which can be equipped with a Riemannian metric, thus giving a meaning to such concepts as "shortest paths" (an approach we will follow ourselves below). See [7].

2.5 Skeletal graphs and Morse theory

This is another shape representation designed to deal with shapes which are not homeomorphic. A shape is represented as a weighted graph, designed so that cycles of vertices in the graph correspond exactly to holes in the original shape. In [4] an algorithm for

designing such a graph is given, but no attempt at comparing them is made. The representation thus designed is inherently invariant under translation, rotation and scaling.

2.6 Point cloud data and the Gromov-Hausdorff distance

A point cloud is a finite set of points $U \subset E$, where E is a metric space which we shall consider a shape (it may be a subset of $\mathbb{R}^n$ or a Riemannian manifold, for example). Then U is a finite metric space, and we may consider the core problem to be that of deciding whether two metric spaces are close to each other or not. This representation is independent of any ambient space, so a regular subspace distance such as the Hausdorff distance,

$$\tilde{d}_Z(X,Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} d_Z(x,y),\ \sup_{y\in Y}\inf_{x\in X} d_Z(x,y)\right\},$$

for $X, Y \subset Z$, will not work. The Gromov-Hausdorff distance between metric spaces,

$$d_{\mathcal{GH}} = \inf_{Z,f,g} \tilde{d}_Z(f(X), g(Y)),$$

where $f : X \to Z$ and $g : Y \to Z$ are isometric embeddings, is explored in [16]. This distance is inherently translation- and rotation-invariant.

2.7 This paper

In this paper, we will use a relatively simple representation of shapes as closed or non-closed curves in $\mathbb{R}^2$ with some strong regularity. The representation we use has been taken from [11] in its entirety: if a curve γ in $\mathbb{R}^2$ is parametrized by arclength, its direction function θ uniquely determines γ up to translation, and by restricting the domain and mean value of θ, we achieve invariance under rotation and scaling as well. This representation is gone through in detail in section 3. In section 4 we state and prove rigorously (more or less, anyway) several results that were left implied or simply brushed over in [11]. In particular, we investigate under which conditions on θ the sets in question really are manifolds. The closed inelastic preshape and shape spaces are found in [11], while the nonclosed spaces are not. The discrete preshape manifold of section 4.4 is also found in [11].

We then present two methods for computing geodesics in the discrete manifold. The shooting method of section 5 is also from [11], but little mention is made there of how the minimization is done. The direct method of section 6 is, to the best of our knowledge, new.

These two methods are compared in a variety of experiments in section 7.

3 Representation

We start by considering a shape as a differentiable parametrized curve $\gamma : [0, 2\pi] \to \mathbb{R}^2$ satisfying $|\gamma'| > 0$. Making the usual identification between $\mathbb{R}^2$ and $\mathbb{C}$, we write

$$\gamma'(t) = \exp[\nu(t) + i\theta(t)].$$

We wish to represent our curve γ in terms of $\nu = \log|\gamma'|$ and the direction function θ. Apart from translation, the actions mentioned above (scaling, rotation and reparametrization) have various easily recognizable effects on the functions ν, θ. Translation, of course, has no effect at all, so our proposed representation is already translation invariant.

3.1 Scaling

We first consider scaling. Scaling a curve has the effect of adding a constant to ν. Customarily, then, we may consider that two pairs of functions (ν_1, θ_1) and (ν_2, θ_2) represent the same shape if $\nu_1 - \nu_2$ is constant. Another option is to strictly enforce some particular ν by requiring that

$$\int_0^{2\pi} e^{\nu(t)}\,\mathrm{d}t = 2\pi,$$

or some other fixed quantity. This approach is taken in [14]. Another choice, which makes the subsequent analysis simpler, is to consider only arclength parametrizations, those where $\nu \equiv 0$, which is done in [11].

Intuitively, fixing $\nu = 0$ renders our shapes inelastic, while allowing ν to vary allows elastic shapes, where the material may "stretch" or "contract" as necessary. The elastic shape spaces might, be better representations of "real" shapes (whatever *that* is). In this paper, we deal only with the inelastic shape spaces, but generalizing the methods to the elastic case might be a worthwile effort for the future.

Thus, for the remainder of the paper, ν will be assumed equal to zero, and the only function representing a shape is θ.

3.2 Rotation

Assuming now that $\nu \equiv 0$, the only remaining freedom is in the direction function θ. We now consider the action of rotation, which has the same effect on θ that scaling had on ν, namely, that of adding a constant. For this reason we will only consider those θ with a vanishing mean,

$$\int_0^{2\pi} \theta(t)\,\mathrm{d}t = 0. \tag{1}$$

We note in passing that in the elastic case, the corresponding requirement should be

$$\int_0^{2\pi} e^{\nu(t)}\theta(t)\,\mathrm{d}t = 0,$$

which, when $\nu = 0$, reduces to (1).

We will often refer to (1) as the *vanishing mean property.*

3.3 Closure and index

We now wish to address the question of closed versus nonclosed shapes. In some cases one may wish to consider only closed shapes (say, if the shapes are to be interpreted as contours). In this case, we assume γ to be periodic and differentiable at the "join" $0 \cong 2\pi$, so that its domain can be interpreted as S^1 instead of $[0, 2\pi]$. Then, the closure conditions must be satisfied, and they are

$$\int_0^{2\pi} \cos\theta(t)\,\mathrm{d}t = \int_0^{2\pi} \sin\theta(t)\,\mathrm{d}t = 0, \tag{2}$$

or, equivalently, the complex formulation

$$\int_0^{2\pi} e^{i\theta(t)}\,\mathrm{d}t = 0. \tag{3}$$

Again, we note that for elastic shapes, the corresponding complex and real formulations are

$$\int_0^{2\pi} e^{\nu(t)+i\theta(t)}\,\mathrm{d}t = 0 \Longrightarrow \int_0^{2\pi} e^{\nu(t)}\cos\theta(t)\,\mathrm{d}t = \int_0^{2\pi} e^{\nu(t)}\sin\theta(t)\,\mathrm{d}t = 0.$$

Closed curves have an *index*, an integer n satisfying $\theta(2\pi) - \theta(0) = 2\pi n$. Let C_n be the set of θ for which the corresponding curve has index n. If we denote by C the set of all θ representing closed curves, the sets C_n form a disconnected partition of C. It is customary to only consider curves in C_1. The reason for this is that curves in C_{-n} are only reparametrizations of curves in C_n, and that for $n = 0$ and $n > 1$, the set C_n consists exclusively of nonsimple curves, which we would usually like to disregard as shapes.

There are, certainly, nonsimple curves in C_1 as well, but there are no easy ways to determine whether a direction function θ represents a curve that is simple or not, and in practice, the effects of nonsimple curves are seldom seen.

3.4 Reparametrization

The issue of reparametrizations remains. In the case of nonclosed curves, this is really a non-issue, because the only possible reparametrization of $\theta(t)$ is the mirrored function $\tau\theta(t) = \theta(2\pi - t)$.[2]

In the case of closed curves, this is not so simple, because any reparametrization of θ of the form $\bar{\theta}(t) = \theta(c + t) - c = c \cdot \theta(t)$ will work, assuming here that θ is extended semiperiodically[3] to all of $\mathbb{R}$. In this case, mirroring will not work, because that yields an index of -1; in restricting our attention to curves with index 1 we have already eliminated reverse parametrizations.

[2]Specifically, it is $\tau\theta(t) = \theta(2\pi - t) + \pi$, but the constant π falls away due to the vanishing mean condition.

[3]We call a function $f : \mathbb{R} \to \mathbb{R}$ *semiperiodic* (with semiperiod $p > 0$ and elevation s) if it satisfies $f(t + np) = f(t) + sn$ for all $n \in \mathbb{Z}$. A semiperiodic function is periodic only if $s = 0$. In our case, $p = s = 2\pi$. We also say that f is (p, s)-semiperiodic. A semiperiodic function is fully determined by the elevation s and its values on one semiperiod, so we can talk of semiperiodic extensions to $\mathbb{R}$. In this paper, we use the term *semiperiodic* to mean $(2\pi, 2\pi)$-semiperiodic.

4 Shape spaces

4.1 The nonclosed inelastic shape space

The nonclosed inelastic shape space consists of, as the name suggests, nonclosed inelastic shapes.

Definition 1. The *nonclosed inelastic preshape space* is

$$\mathcal{C}_{\mathrm{n}} = \left\{ \theta \in L^2(0, 2\pi) \;\middle|\; \int_0^{2\pi} \theta(t)\,\mathrm{d}t = 0 \right\}.$$

The group $\mathbb{Z}_2 = \{\tau, \varepsilon\}$ acts on $\mathcal{C}_{\mathrm{n}}$ with ε as the identity and

$$\tau\theta(t) = \theta(2\pi - t).$$

A preshape $\theta \in \mathcal{C}_{\mathrm{n}}$ is said to be mirror-symmetric if $\tau\theta = \theta$, and the set of all preshapes which are not mirror-symmetric is denoted by $\mathcal{M}_{\mathrm{n}}$. Then the *nonclosed inelastic shape space* is

$$\mathcal{S}_{\mathrm{n}} = \mathcal{M}_{\mathrm{n}}/\mathbb{Z}_2.$$

The following proposition establishes the nature of these spaces.

Proposition 2. *The nonclosed inelastic preshape space $\mathcal{C}_{\mathrm{n}}$ is a linear subspace of the Hilbert space $L^2(0, 2\pi)$, thus a Riemannian manifold. Restricted to shapes that are mirror-asymmetric, $\mathcal{S}_{\mathrm{n}}$ is a quotient manifold of codimension 0. The group $\mathbb{Z}_2$ acts by isometries on $\mathcal{C}_{\mathrm{n}}$, so $\mathcal{S}_{\mathrm{n}}$ becomes a Riemannian manifold in a natural sense.*

Proof. The linearity of $\mathcal{C}_{\mathrm{n}}$ is clear. It inherits the natural inner product from $L^2(0, 2\pi)$. The set of mirror-asymmetric preshapes $\mathcal{M}_{\mathrm{n}}$ is clearly open in $\mathcal{C}_{\mathrm{n}}$, thus a submanifold. Since $\mathbb{Z}_2$ is compact and acts freely on $\mathcal{M}_{\mathrm{n}}$ by definition, $\mathcal{S}_{\mathrm{n}}$ is a quotient manifold restricted to equivalence classes in $\mathcal{M}_{\mathrm{n}}$.

With the notation from definition 13 in the appendix, we have

$$DL_\tau(f) = \tau f,$$

so it's easy to see that $\mathbb{Z}_2$ acts by isometries on $\mathcal{M}_{\mathrm{n}}$. By theorem 14, $\mathcal{S}_{\mathrm{n}}$ has a natural Riemannian structure. □

With these results in mind, the distance metric and the shortest paths on $\mathcal{C}_{\mathrm{n}}$ are easy to construct. Given $\psi, \phi \in \mathcal{C}_{\mathrm{n}}$, the shortest path connecting them in unit time is

$$\theta_s = \psi + s(\phi - \psi), \quad 0 \leq s \leq 1. \tag{4}$$

Since the geometry on $\mathcal{C}_{\mathrm{n}}$ is flat, straight lines, geodesics and shortest paths are exactly the same objects, and moreover, given their boundary points, they are unique. The distance is, not surprisingly perhaps,

$$d_{\mathcal{C}_{\mathrm{n}}}(\psi, \phi) = \|\phi - \psi\| \tag{5}$$

in the $L^2(0, 2\pi)$-norm. Projected down to paths in $\mathcal{S}_\mathrm{n}$, these make sense for the most part. The problem is that even if ψ and ϕ are not mirror-symmetric, $\psi + s(\phi - \psi)$ might be mirror-symmetric for some s. The following proposition limits the occurences of this.

Proposition 3. *Suppose $\psi, \phi \in \mathcal{C}_\mathrm{n}$, and that $\psi + s(\phi - \psi)$ is mirror-symmetric for two distinct times $s = s_1, s_2 \in (0, 1)$. Then $\psi, \phi \notin \mathcal{M}_\mathrm{n}$.*

Proof. The mirror-symmetry means that

$$\begin{aligned} \tau\psi + s_1(\tau\phi - \tau\psi) &= \psi + s_1(\phi - \psi), \\ \tau\psi + s_2(\tau\phi - \tau\psi) &= \psi + s_2(\phi - \psi). \end{aligned}$$

Denoting $R = \psi - \tau\psi$ and $Q = \phi - \tau\phi$, these can be expressed as

$$(s_1 - 1)R = s_1 Q, \qquad (s_2 - 1)R = s_2 Q \qquad \Longrightarrow \qquad \frac{s_1 - 1}{s_1} R = \frac{s_2 - 1}{s_2} R.$$

Since $x \mapsto (x - 1)/x$ is injective on $(0, 1)$, this means $R = 0$ and $Q = 0$, so ψ and ϕ are mirror-symmetric. □

In fact, it is clear that if the geodesic in (4) passes through a mirror-symmetric preshape, almost all perturbations of arbitrary small size to this path yield a path connecting ψ to ϕ that does not become mirror-symmetric. This perturbed path is not a geodesic, but as the perturbations approach zero, the distance the perturbed paths traverse will approach that given in (5). Also, we should not worry that any of the intermediate preshapes become mirror-symmetric: even if $\mathcal{S}_\mathrm{n}$ is not a manifold there, the interpretation of preshapes as curves in $\mathbb{R}^2$ remains: they can still be *drawn*.

In summary, the reduction of (5) to $\mathcal{S}_\mathrm{n}$ gives a shape metric:

$$d_{\mathcal{S}_\mathrm{n}}([\psi], [\phi]) = \min\left\{\|\psi - \phi\|, \|\tau\psi - \phi\|\right\},$$

and the corresponding paths of preshapes are, under projection to $\mathcal{S}_\mathrm{n}$ "almost" geodesic paths of shapes (they might become mirror-symmetric exactly once).

We will now turn our attention to a more interesting space.

4.2 The closed inelastic shape space

Definition 4. The *closed inelastic preshape space* is

$$\mathcal{C}_\mathrm{c} = \left\{ \theta \in S_{sp} \;\middle|\; \int_0^{2\pi} \theta(t)\,\mathrm{d}t = \int_0^{2\pi} \sin\theta(t)\,\mathrm{d}t = \int_0^{2\pi} \cos\theta(t)\,\mathrm{d}t = 0 \right\}. \tag{6}$$

The additive Lie group S^1 acts on S_{sp} according to

$$c \cdot \theta(t) = \theta(c + t) - c. \tag{7}$$

A preshape $\theta \in S_{sp}$ is said to be *rotation-symmetric* if $c \cdot \theta = \theta$ for some $c \neq 0$. The set of all shapes $\theta \in \mathcal{C}_\mathrm{c}$ which are not rotation-symmetric is denoted by $\mathcal{M}_\mathrm{c}$. Then the *closed inelastic shape space* is $\mathcal{S}_\mathrm{c} = \mathcal{M}_\mathrm{c}/S^1$.

The set of rotation-symmetric preshapes is $\mathcal{R}_\mathrm{c} = \mathcal{C}_\mathrm{c} \setminus \mathcal{M}_\mathrm{c}$.

The following proposition establishes some useful results.

Proposition 5. *The closed inelastic preshape space $\mathcal{C}_\mathrm{c}$ inherits a Riemannian metric from S_p, and is thus a Riemannian manifold, of which $\mathcal{M}_\mathrm{c}$ is an open subset, hence a Riemannian submanifold of codimension zero. Restricted to rotation-asymmetric preshapes, $\mathcal{S}_\mathrm{c}$ is a quotient manifold of codmension* 1*. The group S^1 acts by isometries on $\mathcal{C}_\mathrm{c}$, so $\mathcal{S}_\mathrm{c}$ becomes a Riemannian manifold in a natural sense.*

Proof. • *$\mathcal{C}_\mathrm{c}$ is a Riemanniam manifold*: It is a manifold by the level set theorem. The tangent spaces of $\mathcal{C}_\mathrm{c}$ are subspaces of S_p, and inherit its Riemannian metric.

- *$\mathcal{M}_\mathrm{n}$ is an open subset of $\mathcal{C}_\mathrm{c}$*: Let $\theta \in S_{sp}$ be rotation-asymmetric, and consider an arbitrary perturbation ϕ. Then

$$\theta + \epsilon\phi - c \cdot (\theta + \epsilon\phi) = (\theta - c \cdot \theta) + \epsilon\,(\phi - c \cdot \phi)\,.$$

  The function $\theta - c \cdot \theta$ is nonzero by assumption, so for sufficiently small ϵ the entire expression is nonzero. Thus, for such ϵ, $\theta + \epsilon\phi$ is rotation-asymmetric, and this shows that the set of rotation-asymmetric shapes in S_{sp} is open in S_{sp}, hence its intersection with $\mathcal{C}_\mathrm{c}$, which is $\mathcal{M}_\mathrm{c}$, will be open in the subspace topology.

- *The action in (7) is a group action from S^1 on $\mathcal{M}_\mathrm{c}$*: We have

$$c_1 \cdot (c_2 \cdot \theta(t)) = c_1 \cdot (\theta(c_2 + t) - c_2) = \theta(c_1 + c_2 + t) - (c_1 + c_2) = (c_1 + c_2) \cdot \theta(t),$$

  so the action is a group action. We now show that for $\theta \in \mathcal{C}_\mathrm{c}$, $c \cdot \theta \in \mathcal{C}_\mathrm{c}$. Assume $\theta \in \mathcal{C}_\mathrm{c}$, and use the fact that $\theta(t) - t = \tilde{\theta}(t)$ is periodic, to get

$$\begin{aligned}
\int_0^{2\pi} (\theta(c+t) - c)\,\mathrm{d}t &= \int_0^{2\pi} \left(\mathrm{id}_\mathbb{R}(c+t) + \tilde{\theta}(c+t) - c\right) \mathrm{d}t \\
&= \int_0^{2\pi} t\,\mathrm{d}t + \int_0^{2\pi} \tilde{\theta}(c+t)\,\mathrm{d}t \\
&= \int_0^{2\pi} t\,\mathrm{d}t + \int_0^{2\pi} \tilde{\theta}(t)\,\mathrm{d}t \\
&= \int_0^{2\pi} \left(\mathrm{id}_\mathbb{R}(t) + \tilde{\theta}(t)\right) \mathrm{d}t = \int_0^{2\pi} \theta(t)\,\mathrm{d}t = 0,
\end{aligned}$$

  so the vanishing mean property is preserved. For closure, we have

$$\begin{aligned}
\int_0^{2\pi} e^{i(\theta(c+t)-c)}\,\mathrm{d}t &= e^{-ic} \int_0^{2\pi} e^{i\theta(c+t)}\,\mathrm{d}t \\
&= e^{-ic} \int_0^{2\pi} e^{i\theta(t)}\,\mathrm{d}t = 0,
\end{aligned}$$

  since if θ is semiperiodic with period and elevation 2π, then $e^{i\theta}$ is periodic with period 2π. It follows that $c \cdot \theta \in \mathcal{C}_\mathrm{c}$.

The group action also preserves the property of rotation-asymmetry, so that if $\theta \in \mathcal{M}_c$, then $c \cdot \theta \in \mathcal{M}_c$: if $\theta \in \mathcal{C}_c$ is rotation-asymmetric, and $c_1 \in S^1$, so is $c_1 \cdot \theta$. Indeed, suppose $c_1 \cdot \theta$ is rotation-symmetric, so that $(c_2 + c_1) \cdot \theta = c_1 \cdot \theta$ for some $c_2 \in S^1$. Acting with $-c_1$ on the left, we get $c_2 \cdot \theta = \theta$, so θ is rotation-symmetric. Thus the group action is well defined on $\mathcal{M}_c$.

- $\mathcal{S}_c$ *is a quotient manifold of codimension* 1: The group action is free on $\mathcal{M}_c$, and the acting group is compact and of dimension 1. The claim follows.
- S^1 *acts by isometries*: With the notation from definition 13 in the appendix, we have
$$DL_c|_\theta(f) = c \odot f,$$
Since $f \in S_p$ are periodic, it is a breeze to confirm that
$$\langle DL_c|_\theta(f), DL_c|_\theta(g)\rangle_{c\cdot\theta} = \int_0^{2\pi} f(c+t)g(c+t)\,\mathrm{d}t = \int_0^{2\pi} f(t)g(t)\,\mathrm{d}t = \langle f, g\rangle_\theta\,,$$
so S^1 acts by isometries. The rest follows from theorem 14.

□

The next result shows that $\mathcal{R}_c$ is negligible.

Proposition 6. *The set $\mathcal{R}_c$ is a countable disjoint union of affine subspaces of infinite codimension.*[4]

Proof. Let $\theta \in \mathcal{R}_c$. Then it is rotation-symmetric, i.e. there is some nonzero $c \in S^1$, so that $c \cdot \theta = \theta$. Writing $\theta = \mathrm{id}_R + \tilde{\theta}$ with $\tilde{\theta} \in S_p$, we obtain that $c \odot \tilde{\theta} = \tilde{\theta}$.In other words, c is a period of the 2π-periodic function $\tilde{\theta}$, so clearly $\tilde{\theta}$ is constant *or* there is some maximal $n(\theta) \in \mathbb{Z}_+$ so that $\tilde{\theta}$ is $2\pi/n(\theta)$-periodic. By convention, if $\tilde{\theta}$ is constant we write $n(\theta) = 0$.

Denote by $\mathcal{R}_k$ the set of $\theta \in \mathcal{R}_c$ for which $n(\theta) = k$. It is clear that these sets are disjoint and their union is $\mathcal{R}_c$. The first, $\mathcal{R}_0$, is a 1-dimensional affine subspace of S_{sp}, so it has infinite codimension. For $k > 0$, $\mathcal{R}_k$ can be identified with S_p in a natural way, and these are all infinite-dimensional affine subspaces of S_{sp}, with infinite codimensions. □

Finally, a straightforward calculation should verify the next result.

Proposition 7. *The normal space of $\mathcal{C}_c$ at θ is*
$$N_\theta\mathcal{C}_c = \mathrm{span}\{1, \sin\theta, \cos\theta\},$$
and the tangent space is
$$T_\theta\mathcal{C}_c = N_\theta^\perp\mathcal{C}_c = \left\{ f \in S_p \;\middle|\; \int_0^{2\pi} f(t)\,\mathrm{d}t = \int_0^{2\pi} f(t)\sin\theta(t)\,\mathrm{d}t = \int_0^{2\pi} f(t)\cos\theta(t)\,\mathrm{d}t = 0 \right\}.$$

[4] A subspace $U < V$ is of infinite codimension if $\mathrm{codim}_V U > k$ for all $k \in \mathbb{Z}_+$.

4.3 Geodesic paths of closed inelastic shapes

Our ultimate objective is to compute shape distances in $\mathcal{S}_c$. Since $\mathcal{S}_c$ is a Riemannian manifold, the distance function is

$$d_{\mathcal{S}_c}([\psi],[\phi]) = \inf_\gamma L(\gamma),$$

where γ ranges over all piecewise smooth paths connecting $[\theta]$ to $[\phi]$, and

$$L(\gamma) = \int_a^b \|\gamma'(s)\|_{\gamma(s)}\,\mathrm{d}s$$

is the length of the path $\gamma : (a,b) \to \mathcal{S}_c$. If the minimum is achieved for some specific path $\overline{\gamma}$, that path is called a *segment* from $[\psi]$ to $[\phi]$. Segments are shortest paths between their endpoints, but in general, they may not be unique.

It is a known fact that all segments are (if parametrized by arclength) *geodesics*, paths which have no intrinsic acceleration. Geodesics are solutions of a second-order ODE on the manifold in question, an ODE which is determined by the Riemannian metric (see [8] for example). On the other hand, geodesics are, in general, only *locally* segments. That is to say, short pieces of a geodesic are always shortest paths between their endpoints (*unique* shortest paths, actually, if that segment is sufficiently short), whereas this property may not hold for the entire geodesic as a whole. Moreover, singularities may develop which prevents geodesics from extending indefinitely in time and space.

The question of the existence and uniqueness of geodesics is challenging indeed, and more so in infinite dimensions, exemplified, for example, by the fact that the celebrated Hopf-Rinow theorem fails (see [3]). We will return to these questions in section 4.4. In the meantime, we shall assume that we can find shortest paths (or "sufficiently short" paths) by merely computing geodesic paths.

Our ultimate objective is then to calculate geodesic paths on the shape space $\mathcal{S}_c$ with prescribed boundary conditions: given two shapes $[\psi]$ and $[\phi]$, we want to find a path of shapes $[\theta_s]$, $0 \le s \le 1$ connecting them. To achieve this, we will take a detour through geodesic paths of *preshapes* in $\mathcal{C}_c$, a manifold that is easier to work with.

Assuming for a moment that we have a geodesic path θ_s in $\mathcal{C}_c$ whose tangent is normal to the equivalence classes of $\mathcal{S}_c$ wherever they meet, the projected path $\pi(\theta_s)$ (where $\pi : \mathcal{M}_c \to \mathcal{S}_c : \theta \mapsto [\theta]$ is the standard projection) might not be a geodesic in $\mathcal{S}_c$ for the simple reason that θ_s might step outside $\mathcal{M}_c$, i.e. it might become rotation-symmetric. For the same reasons as for the nonclosed case, we take this lightly. The set of rotation-symmetric preshapes in $\mathcal{C}_c$ is negligible indeed, as shown in proposition 6. Since $\operatorname{codim}_{S_{sp}} \mathcal{R}_c = \infty$ and $\operatorname{codim}_{S_{sp}} \mathcal{C}_c = 3$, we must have $\operatorname{codim}_{\mathcal{C}_c}(\mathcal{R}_c \cap \mathcal{C}_c) \ge \infty$ (meaning, $\mathcal{R}_c \cap \mathcal{C}_c$ is a (possibly nonproper) subset of some submanifold of $\mathcal{C}_c$ of infinite codimension). Moreover, almost all small perturbations of the geodesic path should keep us within $\mathcal{M}_c$.

This justifies our planned course of action:

Given shapes $[\psi], [\phi]$ we compute geodesic paths in $\mathcal{C}_c$ connecting ψ to various preshapes in $[\phi]$, the shortest of which, under projection, is "almost" (in the sense just discussed) a shortest path from $[\psi]$ to $[\phi]$ in $\mathcal{S}_c$, and its length is equal to that in $\mathcal{C}_c$.

Remark 8. Why should we not also consider all reparametrizations of ψ? Suppose θ_s is a geodesic connecting $c_1 \cdot \psi$ to $c_2 \cdot \phi$. Then $-c_1 \cdot \theta_s$ is a geodesic of the *same length* connecting ψ to $(c_2 - c_1) \cdot \phi$. Since our ultimate goal is geodesics in $\mathcal{S}_c$, it is helpful to see that, for all $c \in S^1$, the projected geodesic $\pi(c \cdot \theta_s) = [c \cdot \theta_s]$ does not depend on c, by definition. Thus, any geodesic emenating from $c_1 \cdot \psi$ is equivalent, in $\mathcal{S}_c$, to a geodesic emenating from ψ, and vice versa.

The remainder of this paper will deal with strategies for computing paths in $\mathcal{C}_c$ with prescribed boundary conditions.

We will now discretize the manifold $\mathcal{C}_c$, leaving behind a finite-dimensional Riemannian manifold, and argue that geodesics on the discrete manifold are approximations to geodesics on $\mathcal{C}_c$.

4.4 The space of discrete closed inelastic preshapes

Given a $\theta \in \mathcal{C}_c$ we discretize by sampling at N points $t = nh$, $n = 0, \ldots, N-1$, $h = 2\pi/N$, to get the vector $\Theta = (\Theta^0, \ldots, \Theta^{N-1})^T$. By also discretizing the conditions in (6), we get a finite-dimensional manifold, this time embedded in $\mathbb{R}^N$ and of codimension 3, which we shall call $\mathcal{C}_c^N$:

$$\mathcal{C}_c^N = \left\{ \Theta \in \mathbb{R}^N \;\middle|\; \sum_i \Theta^i = \sum_i \sin \Theta^i = \sum_i \cos \Theta^i = 0 \right\}. \tag{8}$$

As in the continuous case, $\mathcal{C}_c^N$ is a manifold by the level set theorem. A straightforward computation will again reveal that the normal space of $\mathcal{C}_c^N$ is a natural discretization of the normal space of $\mathcal{C}_c$ from proposition 7:

$$N_\Theta \mathcal{C}_c^N = \operatorname{span} \{ \mathbb{1}_N, \sin \Theta, \cos \Theta \},$$

where sin and cos are understood to be applied termwise, and

$$T_\Theta \mathcal{C}_c^N = N_\Theta^\perp \mathcal{C}_c^N = \left\{ f \in \mathbb{R}^N \;\middle|\; \sum_i f^i = \sum_i f^i \sin \Theta^i = \sum_i f^i \cos \Theta^i = 0 \right\}.$$

The manifold $\mathcal{C}_c^N$ is also equipped with a scaled version of the natural inner product,

$$\langle f, g \rangle_\Theta = h \sum_i f^i g^i,$$

on $\mathbb{R}^N$. The scaling factor is there to ensure comparability between $\mathcal{C}_c^N$ and $\mathcal{C}_c$, as the metric on $\mathcal{C}_c^N$ is now a natural discretization of the metric on $\mathcal{C}_c$.

We will always use small greek letters θ, ψ, ϕ for continuous preshapes and their corresponding capitals Θ, Ψ, Φ for discrete ones.

The next two propositions establish some important properties of geodesics in $\mathcal{C}_c^N$.

Proposition 9. *Let θ_s for $s \in (0,1)$ be a geodesic in $\mathcal{C}_c$. Pointwise sampling for each s then yields a geodesic Θ_s in $\mathcal{C}_c^N$.*

Proof. Let θ_s, $0 \leq s \leq 1$ be a geodesic in $\mathcal{C}_c$. Thus, $\partial^2\theta_s/\partial s^2 \in N_{\theta_s}\mathcal{C}_c$ for all s. The normal space $N_\theta\mathcal{C}_c$ is spanned by the functions 1, $\cos\theta(t)$ and $\sin\theta(t)$, so an equivalent formulation is that there exists real functions $a^i(s)$, $i = 1,2,3$, so that

$$\frac{\partial^2\theta_s(t)}{\partial s^2} = a^1(s)\sin\theta_s(t) + a^2(s)\cos\theta_s(t) + a^3(s). \tag{9}$$

Write $\Theta_s^i = \theta_s(ih)$. Sampling (9) at times $t = ih$, $i = 0, \ldots, N-1$ yields

$$\frac{\partial^2\Theta_s^i}{\partial s^2} = a^1(s)\sin\Theta_s^i + a^2(s)\cos\Theta_s^i + a^3(s), \tag{10}$$

from which we conclude that $\partial^2\Theta_s/\partial s^2 \in N_{\Theta_s}\mathcal{C}_c^N$, so that Θ_s is a geodesic in $\mathcal{C}_c^N$. □

Proposition 10. *Given any two discrete preshapes $\Psi, \Phi \in \mathcal{C}_c^N$, there exists a geodesic path in $\mathcal{C}_c^N$ connecting them. Moreover, all geodesics on $\mathcal{C}_c^N$ can be extended indefinitely.*

Proof. Since $\mathcal{C}_c^N$ is a closed embedded submanifold of $\mathbb{R}^N$, it is complete in its inner metric. The claim then follows from the Hopf-Rinow theorem (for the original proofs, see [10] for two dimensions and the generalization in [15]). □

We concluded in section 4.3 that the question of existence and uniqueness of geodesics with prescribed boundary conditions is a difficult one on most Riemannian manifolds, let alone infinite dimensional ones.

The corresponding questions for the finite dimensional case are somewhat easier to answer. Proposition 10 ends the question of existence with a postitive answer. Uniqueness is still difficult, however, and we give a short overview of the matters at hand.

For any finite dimensional Riemannian manifold $\mathcal{M}$, there is a smooth positive function $i : \mathcal{M} \to \mathbb{R}$, called the *injectivity radius function*. Given any $x \in \mathcal{M}$, geodesics emenating from x with unit velocity are segments at least up until time $i(x)$, although possibly further (this depends on the initial direction). Equivalently, the exponential map is well defined and injective on the *injectivity ball* $B_{i(x)}(0_x) \subset T_x\mathcal{M}$.[5]

If $\lambda(x) > 0$ is an upper bound for the sectional curvatures at x, we have (see [12])

$$i(x) \geq \frac{\pi}{\sqrt{\lambda(x)}}.$$

[5] We already know from proposition 10 that the exponential map is well defined on *all* of $T_x\mathcal{M}$, but even if $\mathcal{C}_c^N$ were not complete, so that proposition 10 did not apply, it would be well defined on the injectivity ball.

For finite dimensional manifolds, the sectional curvature always has a finite upper bound at any point. If the manifold is compact, the sectional curvature has a positive *global* upper bound $\overline{\lambda}$, in which case the injectivity radius has a positive *global* lower bound $\bar{i}$.

Unfortunately, our manifold is not compact, but there is good reason to believe that there is such a global lower bound for i, still. This is because the three functions determining $\mathcal{C}_{\mathrm{c}}^N$ (8) have $N-1$ periods in the $N-1$-dimensional subspace determined by $\sum_i \Theta^i = 0$, thus the manifold must be periodic and everywhere similar to regions in a precompact subset $P \subset \mathcal{C}_{\mathrm{c}}^N$. Thus, an upper bound of the sectional curvature in P will be an upper bound of the sectional curvature in $\mathcal{C}_{\mathrm{c}}^N$. Since P is precompact, the only reason why such an upper bound may fail to exist is if the curvature goes to infinity somewhere on the boundary. This should not be possible because the three functions from (8) are too regular; in fact they are analytic and Lipschitz continuous.

See for example [5] for an overview of the topic (particularly chapter §6.5).

We now present two methods for computing geodesics in $\mathcal{C}_{\mathrm{c}}^N$. The first is that proposed in [11]. The other is, as far as we are aware, an innovation of this paper.

5 The shooting method

The shooting method is the preferred method of computing geodesics described in [11].

Computing a geodesic with prescribed initial conditions (position and velocity) is easy. It is a first-order ODE in the tangent bundle, which may be integrated using any method. Algorithm 1 shows how to compute such geodesics using a simple Euler method.

Algorithm 1 Computing geodesics in $\mathcal{C}_{\mathrm{c}}^N$ with prescribed initial conditions

Require: $\Theta_0 \in \mathcal{C}_{\mathrm{c}}^N$, $\dot{\Theta}_0 \in T_{\Theta_0}\mathcal{C}_{\mathrm{c}}$, stepsize $k = 1/(M+1)$.

1: **for** $i = 0, 1, \ldots, M$ **do**
2: $\quad \Theta_{i+1} = \sigma_{\Theta_i}(\dot{\Theta}_i)$.
3: $\quad v = P_{\Theta_{i+1}}(\dot{\Theta}_i)$.
4: $\quad \dot{\Theta}_{i+1} = \|\dot{\Theta}_i\| v / \|v\|$.
5: **end for**
6: **return** Θ_i for $i = 0, 1, \ldots, M+1$

Remark 11. In algorithm 1, the function σ_Θ in line 2 is a *retraction*, see definition 15 in the appendix. A retraction provides a local coordinate system in terms of the tangent space that is "correct" to the first order. In this case, a reasonable choice of retraction $\sigma_\Theta(v)$ is to form the vector $\Theta + v \in \mathbb{R}^N$, and move this orthogonally to the level sets of the sums in (8), that is to say, in the subspace spanned by their gradients, until the conditions (8) are satisfied. Algorithm 2 shows how this can be done using Newton's method.

The function P_Θ in line 3 is the orthogonal projection $T_\Theta \mathbb{R}^N \cong \mathbb{R}^N \to T_\Theta \mathcal{C}_{\mathrm{c}}^N$.

The reason for our use of $M+1$ instead of M in algorithm 1 is for compatibility of notation with the direct method, which will be presented in section 6.

Algorithm 2 A numerical retraction to $\mathcal{C}_c^N$

Require: $\Theta \in \mathcal{C}_c^N$, $v \in T_\Theta \mathcal{C}_c^N$, tolerance $\epsilon_r > 0$.

1: $x = \Theta + v$.

2: **while** $\sqrt{h} \max\left\{\left|\sum_i x^i\right|, \left|\sum_i \sin x^i\right|, \left|\sum_i \cos x^i\right|\right\} > \epsilon_r$ **do**

3: Form the jacobian

$$J = \sum_i \begin{pmatrix} 1 & \cos x^i & -\sin x^i \\ \cos x^i & \cos^2 x^i & -\cos x^i \sin x^i \\ -\sin x^i & -\cos x^i \sin x^i & \sin^2 x^i \end{pmatrix}$$

4: and the residual

$$z = \sum_i \left(x^i, \sin x^i, \cos x^i\right)^T.$$

5: Solve the system $Jy = -z$ for $y = (y^1, y^2, y^3)^T$.

6: Update x:

$$x \leftarrow x + y^1 \mathbb{1}_N + y^2 \cos x - y^3 \sin x.$$

7: **end while**

8: **return** x

With that out of the way, we ask to which degree we can use algorithm 1 to compute geodesics with prescribed *boundary* conditions. In [11] this problem is solved by using a shooting method: given two discrete preshapes Ψ and Φ, we search for a $v \in T_\Psi \mathcal{C}_c^N$ such that the geodesic starting at Ψ with initial velocity v ends up as close as possible to Φ in unit time.

In other words, we seek to minimize the functional $E : T_\Psi \mathcal{C}_c^N \to \mathbb{R}$ given by

$$E(v) = \|\exp_\Psi(v) - \Phi\|, \tag{11}$$

where the norm is any reasonably chosen one, possibly unrelated to the Riemannian metric. The exponential mapping $\exp_x(v)$ sends a tangent vector v to the point on the corresponding geodesic a unit time away. In our terms, Θ_{M+1} returned from algorithm 1 with initial conditions Θ_0, $\dot{\Theta}_0$ is a numerical approximation to $\exp_{\Theta_0} \dot{\Theta}_0$.

The problem with this approach is that E does not have a closed form derivative, which prevents us from using standard minimization techniques such as steepest descent. A numerical approximation of $\nabla E(v)$ can be computed, but this greatly increases the computational load, as we must calculate $O(N)$ initial-condition geodesics for each iteration (recall that $\dim T_\Psi \mathcal{C}_c^N = N - 3$). Nevertheless, we see no other alternative.

It remains to detail how we approach the equivalence-class problem: in the greater picture, we do not merely seek a geodesic connecting two preshapes, but two equivalence classes of such. We make two modifications to deal with this:

1. The geodesics in $\mathcal{C}_c$ should also be normal to the orbits $[\theta]$ in $\mathcal{C}_c$. These orbits are 1-dimensional, and their tangent space is $T_\theta[\theta] = \text{span}\{\theta' - 1\}$. In $\mathcal{C}_c^N$, we

can approximate the action from S^1 to $\mathcal{C}_c$ by an action from $\mathbb{Z}_N$ to $\mathcal{C}_c^N$, which essentially shifts vectors:

$$(n \cdot \Theta)^i = \Theta^{i+n} - hn,$$

where Θ is assumed to be extended semiperiodically so that it is defined for all $i \in \mathbb{Z}$.[6] These orbits are discrete, however, and do not have a nontrivial tangent space. We work around this problem by explicitly forcing our geodesics to be normal also to the tangent vector

$$\Theta' = P_\Theta \left(\frac{1}{h}(1 \cdot \Theta - \Theta) - \mathbb{1}_N \right).$$

We denote by $\overline{T}_\Theta \mathcal{C}_c^N$ the restricted tangent space. We achieve the desired result by replacing, in algorithm 1, line 3, $P_{\Theta_{i+1}}$ with $\overline{P}_{\Theta_{i+1}}$ where $\overline{P}_\Theta$ is the orthogonal projection $T_\Theta \mathbb{R}^N \cong \mathbb{R}^N \to \overline{T}_\Theta \mathcal{C}_c^N$.

2. After approximating $\exp_\Theta(v)$ using algorithm 1 we instead use the functional F defined by

$$F(v) = \min_{n \in \mathbb{Z}_n} \| \exp_\Theta(v) - n \cdot \Phi \|.$$

This minimization is actually very fast, thanks to an elegant algorithm based on the fast fourier transform, which is detailed in [13].

To summarize, the entire method is given in algorithm 3.

Algorithm 3 Approximating geodesics in between equivalence classes in $\mathcal{C}_c^N$ using the shooting method

Require: $\Psi, \Phi \in \mathcal{C}_c^N$, $v \in \overline{T}_\Psi \mathcal{C}_c^N$, stepsize $k = 1/(M+1)$, tolerance $\epsilon > 0$.

1: Compute an approximation $\tilde{\Phi}$ to $\exp_\Psi(v)$ using algorithm 1 with the described modifications.
2: Use the algorithm in [13] to compute $F = \min_{n \in \mathbb{Z}_n} \|\tilde{\Phi} - n \cdot \Phi\|$.
3: **while** $F > \epsilon$ **do**
4: Compute $g \approx -\nabla F(v)$ by numerical differentiation.
5: Compute some stepsize b using a line-search method.
6: $v = v + bg$.
7: Execute lines 1 and 2.
8: **end while**
9: **return** v.

Remark 12. In line 4, the numerical differentiation requires a basis of sorts for $\overline{T}_\Theta \mathcal{C}_c^N$. We write, as in [11], the vector v as the projection of a truncated Fourier series, and then minimize in terms of the coefficients of this series.

In line 5, one could use for example Armijo's method (see for example [1]). The parameters in this method should be tweaked for optimal performance.

[6] We say that a function $f : \mathbb{Z} \to \mathbb{R}$ is semiperiodic with period N and elevation Nh if $f(i+N) = f(i) + Nh$ for all $i \in \mathbb{Z}$. In the example, we must have $Nh = 2\pi$ for this to make sense as a discretization of the continuous case.

After the algorithm has converged (*if* it has converged) the geodesic can be reconstructed from v using algorithm 1, and $\|v\|_\Psi$ is an approximation to $d_{\mathcal{S}_c}([\psi],[\phi])$.

6 A direct method

A path in $\mathcal{C}_c^N$ is geodesic if and only if the acceleration is always normal to $\mathcal{C}_c^N$ in $\mathbb{R}^N$. In the infinite-dimensional case, this is contained in the proof of proposition 9, and in particular in (9). The finite-dimensional case is quite similar, thanks to the proof of proposition 9.

We propose a more direct approach than the shooting method, based on the criteria in (10). For geodesics in $\mathcal{C}_c^N$, we seek real functions Θ_s^i for $0 \le s \le 1$ and $i = 0, \dots, N-1$, smooth in s, with $\Theta_0 = \Psi, \Theta_1 = \Phi$ given, and real functions $a^1(s), a^2(s), a^3(s)$, so that

1. the path is geodesic, i.e. for all $0 < s < 1$ and all i,

$$\frac{\partial^2 \Theta_s^i}{\partial s^2} = a^1(s) \sin \Theta_s^i + a^2(s) \cos \Theta_s^i + a^3(s)$$

2. the path lies in $\mathcal{C}_c$, i.e. for all $0 < s < 1$,

$$\sum_i \Theta_s^i = \sum_i \sin \Theta_s^i = \sum_i \cos \Theta_s^i = 0.$$

This problem is reminiscent of ODEs that are often successfully solved using a simple difference scheme, which we now utilize.

We can discretize this problem using familiar tecniques. We introduce a grid with stepsize $k = \frac{1}{M+1}$ in the s-direction, and we designate our approximations by $\Theta_{(m)}^n \approx \Theta_{mk}^n$ and $A_m^i \approx a^i(mk)$ for $i = 1, 2, 3$, $m = 1, \dots, M$. In the following we denote by $\Theta_{(m)}$ the vector $(\Theta_{(m)}^0, \dots, \Theta_{(m)}^{N-1})^T$. The two conditions from above then become, for $m = 1, \dots, M$,

$$\begin{aligned}
F_m &\doteq \frac{1}{k^2}(\Theta_{(m+1)} - 2\Theta_{(m)} + \Theta_{(m-1)}) - A_m^1 \sin \Theta_{(m)} - A_m^2 \cos \Theta_{(m)} - A_m^3 = 0, \\
H_m^1 &\doteq h \sum_{n=0}^{N-1} \sin \Theta_{(m)}^n = 0, \\
H_m^2 &\doteq h \sum_{n=0}^{N-1} \cos \Theta_{(m)}^n = 0, \\
H_m^3 &\doteq h \sum_{n=0}^{N-1} \Theta_{(m)}^n = 0,
\end{aligned}$$

where we define $\Theta_{(0)} = \Psi$ and $\Theta_{(M+1)} = \Phi$.

By collecting the unknown quantities into a large vector,

$$\hat{\Theta} = \begin{pmatrix} \Theta_{(1)} \\ \Theta_{(2)} \\ \vdots \\ \Theta_{(M)} \end{pmatrix}, \quad A^i = \begin{pmatrix} A^i_1 \\ A^i_2 \\ \vdots \\ A^i_M \end{pmatrix}, \quad X = \begin{pmatrix} \hat{\Theta} \\ A^1 \\ A^2 \\ A^3 \end{pmatrix},$$

and the depending functions likewise,

$$F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_M \end{pmatrix}, \quad H^i = \begin{pmatrix} H^i_1 \\ H^i_2 \\ \vdots \\ H^i_M \end{pmatrix}, \quad Q = \begin{pmatrix} F \\ H^1 \\ H^2 \\ H^3 \end{pmatrix},$$

we can formulate the discretized problem as finding an X so that $Q(X) = 0$. This can be done using Newton's method, because in fact, the Jacobian of Q is not difficult to calculate. In block form, it is

$$\frac{\partial Q}{\partial X} = \begin{pmatrix} \frac{\partial F}{\partial \hat{\Theta}} & \frac{\partial F}{\partial A^1} & \frac{\partial F}{\partial A^2} & \frac{\partial F}{\partial A^3} \\ \frac{\partial H^1}{\partial \hat{\Theta}} & & & \\ \frac{\partial H^2}{\partial \hat{\Theta}} & & & \\ \frac{\partial H^3}{\partial \hat{\Theta}} & & & \end{pmatrix}.$$

The blocks are given according to the following. Here, $\hat{\Theta}_i$ denotes a single element of the large vector $\hat{\Theta}$, and not a subvector.

- $\partial F/\partial \hat{\Theta}$ is $MN \times MN$ with

$$\left(\frac{\partial F}{\partial \hat{\Theta}}\right)_{i,i} = -\frac{2}{k^2} - A^1_{\lceil i/N \rceil} \cos \hat{\Theta}_i + A^2_{\lceil i/N \rceil} \sin \hat{\Theta}_i, \qquad \left(\frac{\partial F}{\partial \hat{\Theta}}\right)_{i,i\pm N} = \frac{1}{k^2}.$$

- $\partial F/\partial A^j$ are all $MN \times M$ with

$$\left(\frac{\partial F}{\partial A^j}\right)_{i,\lceil i/N \rceil} = \begin{cases} -\sin \hat{\Theta}_i, & j = 1, \\ -\cos \hat{\Theta}_i, & j = 2, \\ -1, & j = 3. \end{cases}$$

- $\partial H^j/\partial \hat{\Theta}$ are all $M \times MN$ satisfying

$$\frac{\partial H^1}{\partial \hat{\Theta}} = -h \frac{\partial F}{\partial A^2}^T, \qquad \frac{\partial H^2}{\partial \hat{\Theta}} = h \frac{\partial F}{\partial A^1}^T, \qquad \frac{\partial H^3}{\partial \hat{\Theta}} = -h \frac{\partial F}{\partial A^3}^T.$$

We have now just about everything we need to implement a Newton's Method algorithm for finding a zero of Q. The only things left to address are firstly, the problem of finding a good initial guess and secondly, how to solve the corresponding linear system. We address these in turn.

An initial guess can be constructed from the inital and final discrete preshapes Ψ, Φ by forming the line $\tilde{\Theta}_s = \Psi + s(\Phi - \Psi)$ in $\mathbb{R}^N$, sampling these at the appropriate s-values ($s = km$ for $m = 1, \ldots, M$), and projecting these back onto $\mathcal{C}_{\mathrm{c}}^N$ using algorithm 2 (with $x = \tilde{\Theta}_{km}$ as the first iterate). This gives us the values $\Theta_{(m)}^n$. The values A_m^i are then computed by projecting the discrete accelerations $(\Theta_{(m+1)} - 2\Theta_{(m)} + \Theta_{(m-1)})/k^2$ onto the space $\operatorname{span}\{\sin\Theta_{(m)}, \cos\Theta_{(m)}, 1\}$.

For each iteration with Newton's method we need to solve a linear system

$$\frac{\partial Q}{\partial X} Z = -Q(X),$$

after which the next iterate is $X + Z$. The matrix $\partial Q/\partial X$ has some structure which can be exploited. We introduce the following partitioning of the matrix $\partial Q/\partial X$ and the vectors Z and $-Q(X)$:

$$\frac{\partial Q}{\partial X} = \begin{pmatrix} A & C \\ D^T & \end{pmatrix}, \qquad Z = \begin{pmatrix} c \\ d \end{pmatrix}, \qquad -Q(X) = \begin{pmatrix} x \\ y \end{pmatrix},$$

where A is $\partial F/\partial\Theta$, C contains the derivatives of F with respect to the a^i, and D^T contains the derivatives of H^i with respect to Θ. This means that A is $MN \times MN$, C and D are $MN \times 3M$, c and x are column MN-vectors, and d and y are column $3M$-vectors. The matrices C and D are related through

$$D = CW, \qquad W = \begin{pmatrix} -hI_M & & \\ & hI_M & \\ & & -hI_M \end{pmatrix}.$$

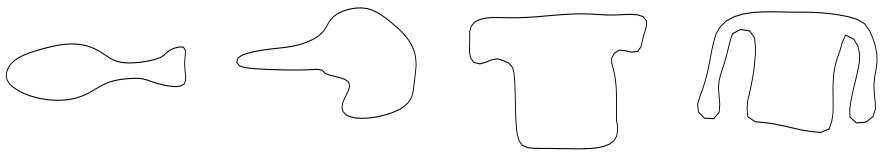
It can be checked that the solution is given by

$$\begin{aligned} (C^T A^{-1} C)d &= W^{-1}\left(D^T A^{-1} x - y\right), \\ Ac &= x - Cd. \end{aligned}$$

If a good method for inverting $A = \partial F/\partial\Theta$ can be found, this might be exploited, because then the solution of the last system will be given through simple multiplication with A^{-1}, whereas the first system can be solved faster with conventional methods (it is of size $O(M)$ instead of $O(MN)$.

Indeed, observe that $A = T/k^2 + S$, where T is the discrete one-dimensional Laplacian, which is independent of X, and S is a diagonal matrix depending on X. We have

$$k^2 T^{-1} A = k^2 T^{-1}\left(\frac{1}{k^2} T + S\right) = I + k^2 T^{-1} S = I + O(k^2),$$



(a) The fish. (b) The duck. (c) The t-shirt. (d) The sweater.

Figure 1: The numerical labrats.

Since $\|T^{-1}\|$ is $O(1)$ and it seems very reasonable that S also is. Thus, k^2T^{-1} is a good approximation to A^{-1} that does not change from iteration to iteration.

As we will see in section 7.2, it may be desirable to finish off with a number of iterations without this approximation, i.e. *after* the quasi-Newton method has converged to a sufficient level of error, using k^2T^{-1} for A^{-1} as above, we perform a number of iterations with the full Newton method, using *no* approximations. We denote the number of such iterations to be β, a nonnegative integer. We will see in section 7 that relevant values are $\beta = 0, 1, 2$. In this way β becomes a parameter.

The workings of this method is wrapped up completely in section 7.3.

7 Results and comparison

7.1 Implementation details

All the code has been executed on an Intel Core2 Duo T9300 processor at 2.50 GHz with an 800 MHz front side bus and 6 MB cache, in an Octave 3.0.1 environment.

We implemented the shooting method with a steepest descent approach using numerical differentiation for approximating the gradients. The auxiliary norm in (11) used to measure closeness was taken to be a rescaled 2-norm.

We have primarily used four shapes for testing purposes. They can be seen in figure 1.

7.2 Measuring geodesicity

Since a geodesic is, by definition, a path with zero intrinsic acceleration, it makes sense to measure geodesicity by computing an approximation to the acceleration in each point, project it down to the tangent space and taking the norm, subsequently averaging over all points along the path.

In the shooting method, the geodecisity condition is implemented through Euler's method, which is of order 1, while in our direct method it is implemented as a condition of order 2, both in $k = 1/(M + 1)$. For comparison purposes, therefore, we use an order 4 approximation of the acceleration to measure geodesicity.

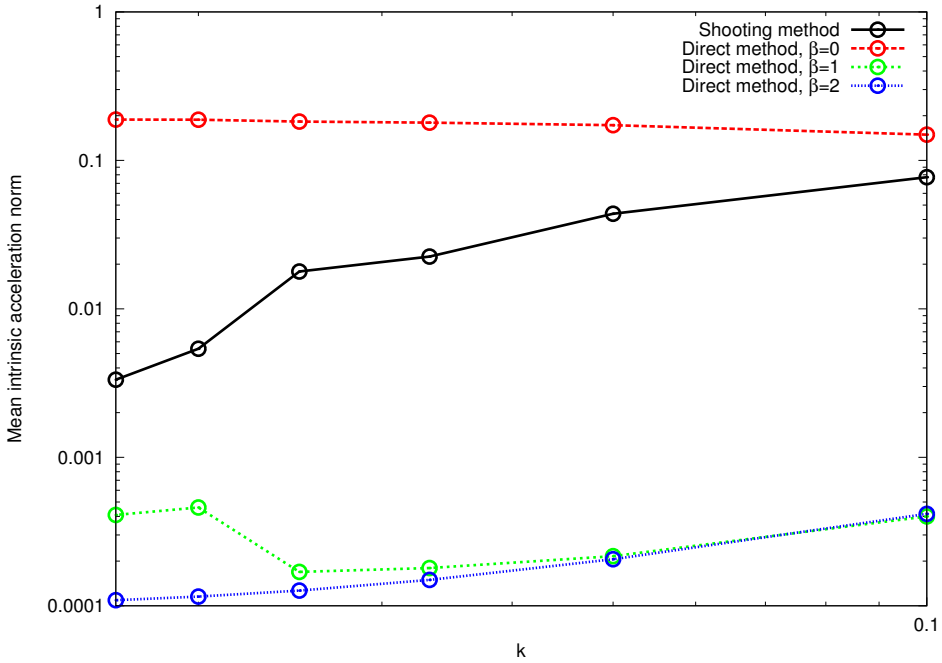


Figure 2: Plot of mean intrinsic acceleration norm vs. $k = 1/(M+1)$ for the fish $\rightarrow$ duck geodesic, for the shooting method and direct method, using $\beta = 0, 1, 2$.

In figure 2, a plot of mean intrinsic acceleration vector vs. k can be found. It shows that the results for the direct method vary with the parameter β. For $\beta = 0$, the shooting method produces geodesics of higher geodesicity, whereas for $\beta > 0$, the direct method is better. Moreover, we can see that the order of convergence in k for the shooting method is close to 1, as one can expect, while for the direct method it is decidedly not 2, in fact it seems less than 1! This might be related to the fact that geodesicity is not *strictly* enforced by a second-order condition, rather it is allowed only to converge to a sufficiently small level of tolerance.

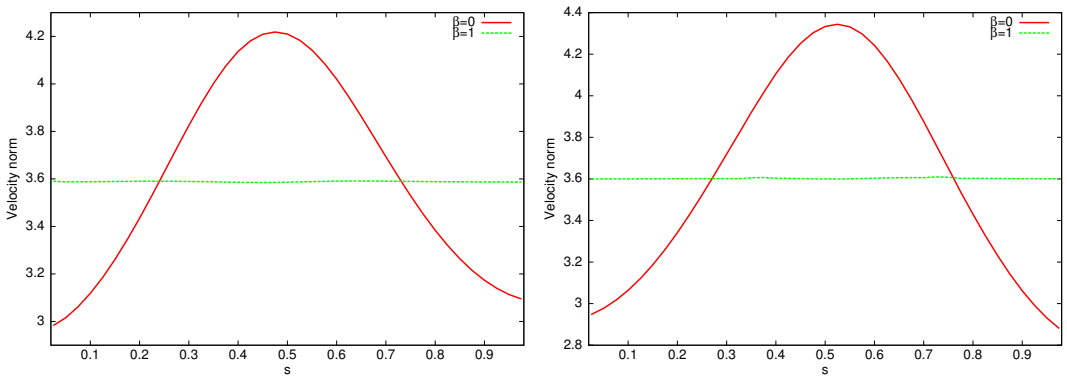
It might be possible that if we allow N to vary together with M, we might achieve a higher order of geodesicity in the direct method. We have not explored this here, however.

7.3 Measuring distance

As mentioned, given a geodesic path Θ_s, connecting Ψ to Φ, the distance between the two is given by

$$d_{\mathcal{C}_c^N}(\Psi, \Phi) = \|\partial\Theta_s/\partial s\|_{\Theta_s},$$

which, by geodesicity, is constant with respect to s. As such, in the direct method, the velocity may be sampled at any point along the geodesic, or averaged over the whole. The discrete geodesics are only approximations, however, and we cannot expect the



(a) T-shirt → sweater.

(b) Sweater → fish.

Figure 3: Velocity norm vs. s for some geodesics with the direct method, for $\beta = 0, 1$.

velocity vector to remain constant in norm.

Figure 3 shows some plots of the velocity norm vs. s for some geodesics with the direct method, for $\beta = 0, 1$. The plots show that the velocity does indeed vary systematically, albeit quite little for $\beta \geq 1$ (on the order of 10^{-2} at most). For $\beta = 0$, the velocity varies a lot, but it would appear from the plots that the mean value does not vary with β. We will see better evidence for this in figure 5. In any case, evidence is that we should take the *mean* velocity if we should have any hope for accuracy with $\beta = 0$.

The minimization alluded to in section 4.3 manifests itself in the direct method, in that we need to compute N different geodesics, corresponding to the N different $i \cdot \Phi$. We cut down the number of different cases to check to $O(\sqrt{N})$ with a simple heuristic method. Instead of computing directly

$$\tilde{d} = \inf_n d_{\mathcal{C}_c^N}(\Psi, n \cdot \Phi),$$

we first compute

$$n_0 = \arg \inf_{n=k\lceil\sqrt{N}\rceil} d_{\mathcal{C}_c^N}(\Psi, n \cdot \Phi),$$

over $k = 1, \ldots, \lceil\sqrt{N}\rceil$, and then

$$\tilde{d} = \inf_{n=n_0+k} d_{\mathcal{C}_c^N}(\Psi, n \cdot \Phi),$$

over $k = -\lceil\sqrt{N}/2\rceil, \ldots, \lceil\sqrt{N}/2\rceil$.

In the shooting method the computation of distance is largely a non-issue, because the velocity vector, which exists explicitly in the algorithm, is also explicitly kept at constant norm throughout. The minimization, too, can be done very efficiently as already explained.

7.4 Geodesics and their lengths

In figure 4 we have plotted the lengths of geodesics $\Psi \to i \cdot \Psi$ as a function of $c = ih \in S^1$, where the preshapes Ψ, Φ are the fish, duck, t-shirt and sweater. This shows that the heuristic minimization procedure in section 7.3 performs well on shapes this smooth.

A logarithmic plot with distance vs. mean intrinsic acceleration norm can be found in figure 5. From this, we note the following:

1. Even though paths computed with $\beta = 0$ have a much higher mean intrinsic acceleration norm than those for $\beta > 0$, the computed length seems to be *constant* with respect to β.

2. The paths found by the shooting method are consistently longer than those found by the direct method.

7.5 Error tolerances

To ensure a fair comparison between the two methods, we should adjust the error tolerance levels, so that they coincide. We have decided to use the infinity norm for this, well suited as it is for our purposes.

There are two conditions to consider. First, that the points on our path are sufficiently close to the manifold $\mathcal{C}_c^N$. Fairness here is easy to enforce, because this condition is enforced in the same way in both methods. Given an $\epsilon > 0$ we require that $\Theta_{(m)}$, $m = 1, \ldots, M$, satisfy

$$\max\left\{\left|h\sum_{n=0}^{N-1}\Theta_{(m)}^n\right|, \left|h\sum_{n=0}^{N-1}\sin\Theta_{(m)}^n\right| \left|h\sum_{n=0}^{N-1}\cos\Theta_{(m)}^n\right|\right\} \leq \epsilon. \tag{12}$$

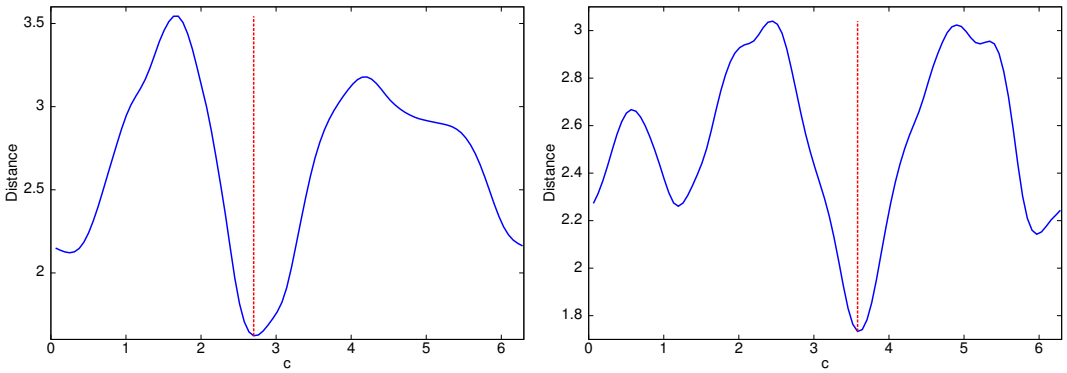
These conditions can be enforced in the convergence requirements for the Newton iteration in the direct method, and in the convergence requirements for the retraction in the shooting method (algorithm 2).

The other condition is that of geodesicity. There seems to be no easy way to ensure "fairness" here, since, as already mentioned, in the shooting method this is enforced explicitly with an order 1 method, while in the direct method it is enforced differently. We rest our case on the previous analysis, in that the geodesicity is comparable, and indeed, the direct method produces better results when $\beta \geq 1$.

7.6 Time consumption and robustness

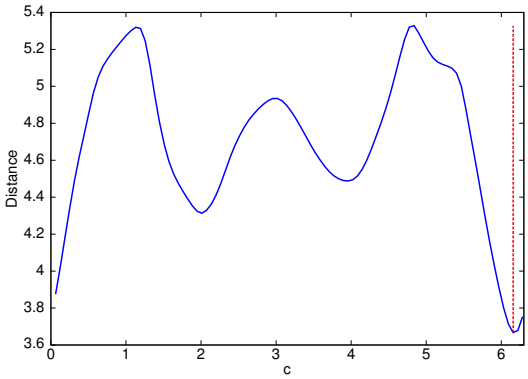
The time consumption for computing some geodesics can be found in table 1. It shows a steady factor of improvement ranging from $85\% - 99\%$, depending on M and β, which can only be described as good.

It is noticeable that the direct method is much more stable in its time consumption, with (estimated) relative standard deviations of $1\% - 7\%$ where the shooting method lies around 70%.

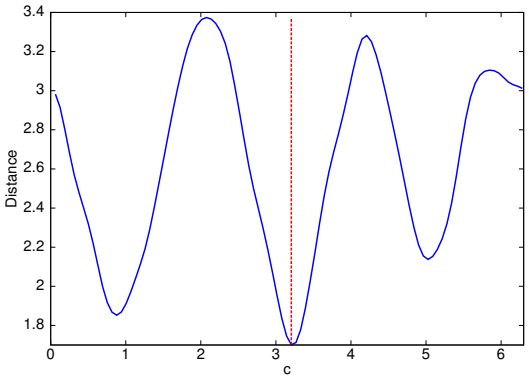


(a) Fish → duck.

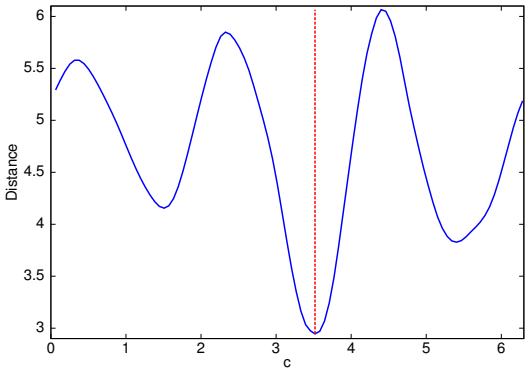
(b) Fish → t-shirt.



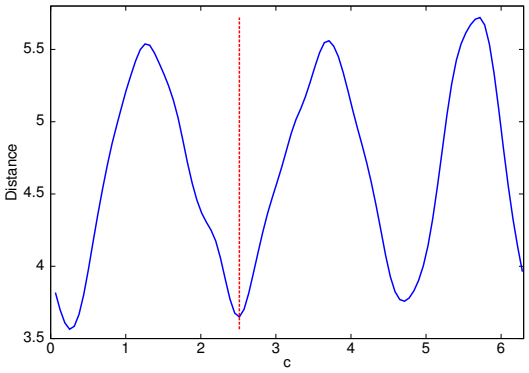
(c) Fish → sweater.



(d) Duck → t-shirt.



(e) Duck → sweater.



(f) T-shirt → sweater.

Figure 4: Lengths of geodesics in $\mathcal{C}_c$ vs. the reparametrization parameter $c \in S^1$. The vertical red line indicates the minimum found by the heuristic algorithm described in section 7.3.

$M = 10$							
Geodesic	Shooting	Direct ($\beta = 0$)		Direct ($\beta = 1$)		Direct ($\beta = 2$)	
F→D	19.7	0.694	96.5%	0.928	95.2%	1.12	94.3%
F→T	23.3	0.685	97.1%	0.867	96.3%	1.15	95.1%
F→S	38.7	0.772	98.0%	0.936	97.6%	1.22	96.8%
D→T	33.4	0.631	98.1%	0.872	97.4%	1.13	96.6%
D→S	96.0	0.676	99.3%	0.890	99.1%	1.13	98.8%
T→S	58.0	0.679	98.8%	0.892	98.5%	1.21	97.9%
RSD	63%	6.7%		3.2%		3.8%	
$M = 20$							
Geodesic	Shooting	Direct ($\beta = 0$)		Direct ($\beta = 1$)		Direct ($\beta = 2$)	
F→D	30.1	1.35	95.5%	2.64	91.2%	4.36	85.6%
F→T	38.0	1.34	96.5%	2.64	93.1%	4.08	89.3%
F→S	47.4	1.41	97.0%	2.73	94.2%	4.22	91.1%
D→T	38.5	1.35	96.5%	2.72	92.9%	4.18	89.1%
D→S	169	1.37	99.2%	2.73	98.4%	4.17	97.5%
T→S	86.6	1.38	98.4%	2.79	96.8%	4.24	95.1%
RSD	78%	1.9%		2.2%		2.2%	
$M = 40$							
Geodesic	Shooting	Direct ($\beta = 0$)		Direct ($\beta = 1$)		Direct ($\beta = 2$)	
F→D	70.3	3.12	95.6%	6.36	91.0%	8.87	87.4%
F→T	56.8	2.34	95.7%	5.55	90.2%	8.93	84.3%
F→S	112	2.40	97.9%	5.80	94.8%	9.11	91.9%
D→T	97.8	2.26	97.7%	5.55	94.3%	8.88	90.9%
D→S	268	2.36	99.1%	5.67	97.9%	9.06	96.7%
T→S	150	2.37	98.4%	5.73	96.2%	9.88	93.4%
RSD	61%	13%		5.2%		4.2%	

Table 1: Execution times of the shooting method and the direct method for $\beta = 0, 1, 2$ and $M = 10, 20, 40$. The factors of improvement are listed besides the times in percent. The sample relative standard deviation (RSD) is also given. The t-resolution was $N = 100$, and in (12) we used $\epsilon = 10^{-2}$. All times are in seconds.

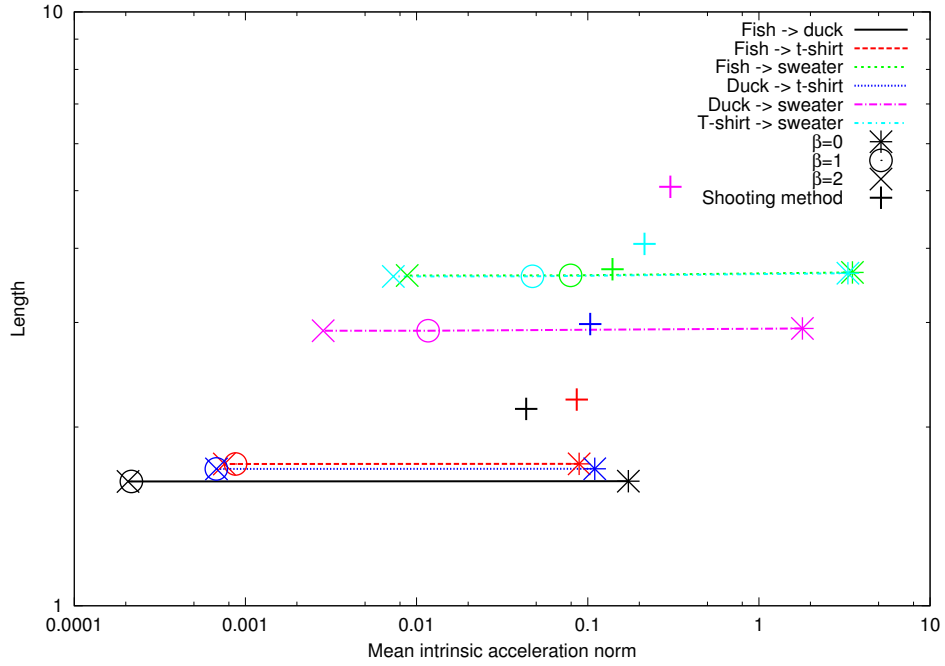


Figure 5: Distance vs. mean intrinsic acceleration norm for the six geodesics, using the two different methods and various choices of β for the direct method.

The asymptotic time complexity of *one iteration* of the shooting method is $O(MN)$, whereas the direct method, as we have implemented it, requires $O(M^2N^2 + M^3)$, disregarding the initial inversion of the matrix T.

The six geodesics, computed at $M = 20$ are shown in figures 6 (for the shooting metod) and 7 (for the direct method), sampled at six uniform points in the s-domain. Some artificial rotation has been imposed, varying with s, to ensure that both the initial and final shapes are rotated correctly. We remind the reader that the shape representation is independent of rotation and that the geodesics could very well not rotate whatsoever. From the figures, we note a few things.

First, that there is, in general, very little agreement between the two methods. Only the fish $\rightarrow$ sweater geodesics in figures 6c and 7c seem to correspond closely. In all other cases the two methods seem to converge to different geodesics. It would be interesting to know whether these geodesics in $\mathcal{C}_c$, $\mathcal{C}_c^N$ or $\mathcal{S}_c$ are unique, given their boundary values, and if so, which of these geodesics are "mirages", and which are "true".

Second, that in cases where the initial and final shapes are (nearly) mirror-symmetric, the shooting method tends to take us through intermediate shapes which are not (figures 6b and 6f). This certainly seems contrary to intuition. The shooting method seems to preserve mirror-symmetry to a higher degree (figures 7b, 7c and 7f), and the only case in which the shooting method preserves mirror-symmetry (figure 6c), is the case where it agrees with the direct method.

Thirdly, we note that in our opinion (highly subjective, no doubt), the geodesics produced by the direct method seem arguably more natural and simple than those produced by the shooting method.

7.7 Problematic shapes

In table 1, we note that the shooting method converges noticeably slower in two cases marked with red, both involving the sweater (the exception is the fish → sweater case, but even this takes very long when M gets large). Is this a coincidence? Why is the sweater so problematic?

7.7.1 Smoothness

The answer, we conjecture, is that θ_{S} (the supposed infinite-dimensional analogue of our finite-dimensional representation Θ_{S} of the sweater shape) is less smooth than the others, in a sense that will now be made more precise.

We recall that in proposition 5, we had to assume smoothness of θ so that the equivalence classes in $\mathcal{C}_{\mathrm{c}}$ became embedded submanifolds. This was required to make the construct $\mathcal{S}_{\mathrm{c}}$, and geodesics on it, meaningful. It should come as no big surprise that the regularity of θ might influence the performance of our algorithms, in particular the speed of convergence of the shooting method, and the geodesicity of the resulting geodesics. In fact, in section 5, we had to numerically approximate θ', so that our geodesics could be forced to run (approximately) normal to the equivalence classes. This approximation was of first order, and the leading error term is proportional to θ''. This indicates that a measure of magnitude of θ'' might provide a decent measure of "problematicity".

An examination of figure 8, which shows plots of θ and its derivative θ' for the four shapes, will reveal that the sweater achieves the highest values of $|\theta''|$. Indeed, their L^2-norms are (using subscripts F, D, T, S for fish, duck, t-shirt, sweater respectively):

$$\begin{aligned} \|\theta''_{\mathrm{F}}\| &= 86.00, & \|\theta''_{\mathrm{D}}\| &= 108.8, \\ \|\theta''_{\mathrm{T}}\| &= 118.2, & \|\theta''_{\mathrm{S}}\| &= 182.2. \end{aligned}$$

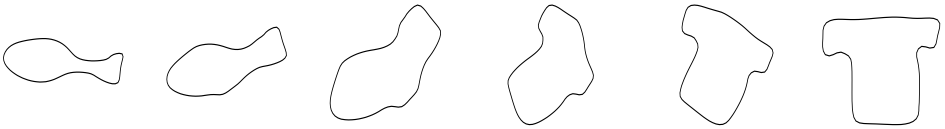
Not surprisingly, perhaps, the sweater achieves easily the highest value.

To test this hypothesis, we used as test shapes the functions

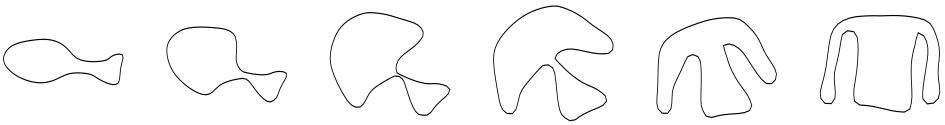
$$\theta_l(t) = t - \sin(2lt) - \pi,$$

for $l \in \mathbb{Z}_+$, see figure 9. One can check that $\theta_l \in \mathcal{C}_{\mathrm{c}}$, and that $\|\theta''_l\| = O(l^2)$. Note that these θ_l are rotation-symmetric. In any case, experiments using these showed in cases involving shapes such as the duck or the fish, and one of the θ_l, the shooting method converged *faster* if l was large, usually with only one iteration if $l \geq 5$. It would seem that the conjecture is false, or at least that there is more to the smoothness matter than we have suspected.

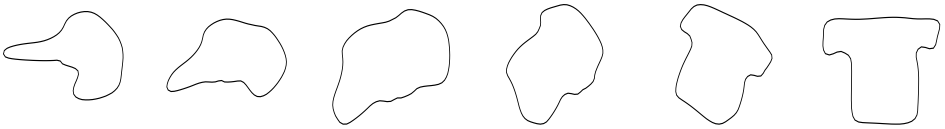
(a) Fish → duck.



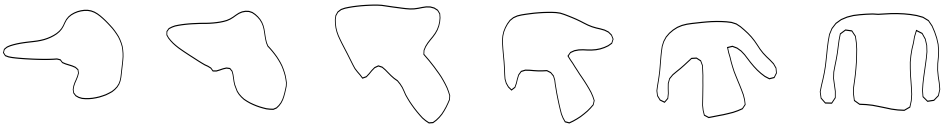
(b) Fish → t-shirt.



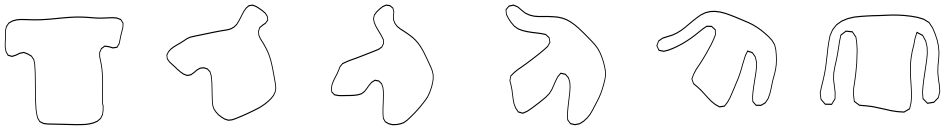
(c) Fish → sweater.



(d) Duck → t-shirt.



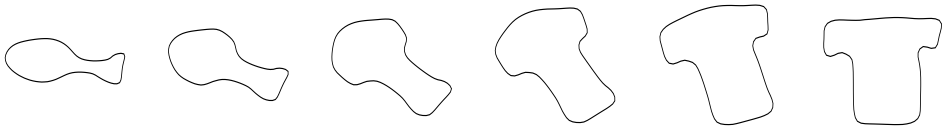
(e) Duck → sweater.



(f) T-shirt → sweater.

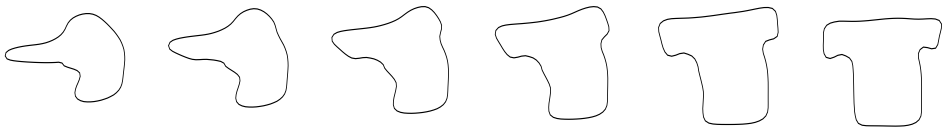
Figure 6: Geodesics with the shooting method.

(a) Fish → duck.



(b) Fish → t-shirt.

(c) Fish → sweater.

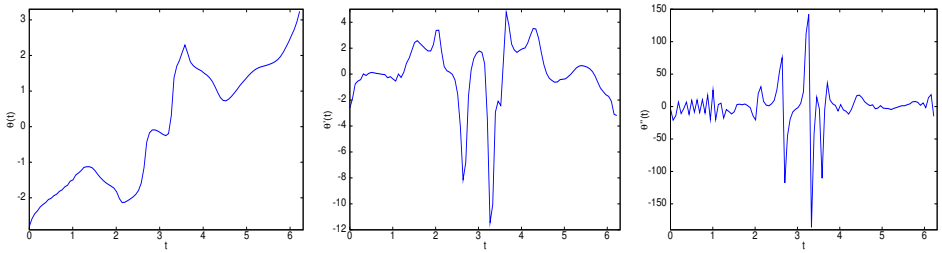


(d) Duck → t-shirt.

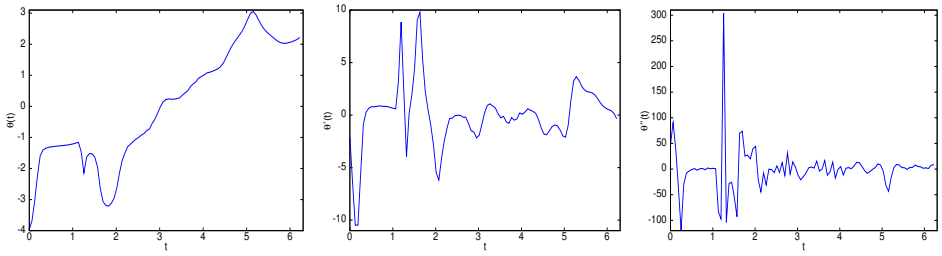
(e) Duck → sweater.

(f) T-shirt → sweater.

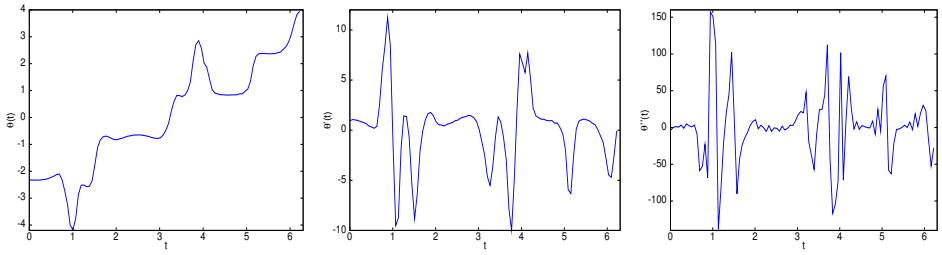
Figure 7: Geodesics with the direct method.



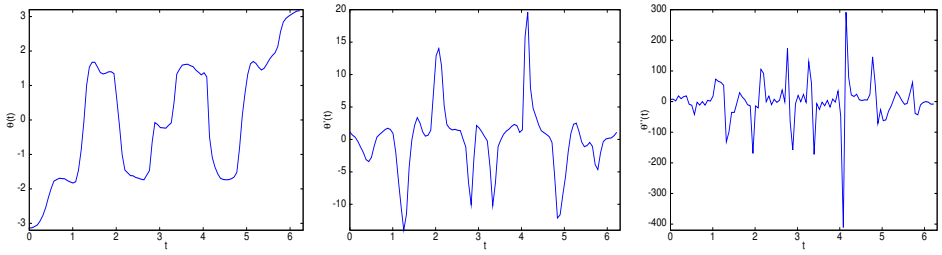
(a) Fish



(b) Duck

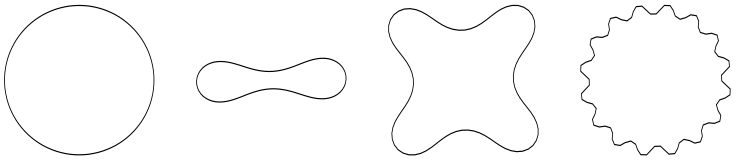


(c) T-shirt



(d) Sweater

Figure 8: Plots of θ (left), θ' (middle) and θ'' (right) for the four shapes. Note the scales.



(a) θ_0 (b) θ_1 (c) θ_2 (d) θ_8

Figure 9: The shapes used to test the theory of smoothness. As l grows, the shapes oscillate more.

7.7.2 Length

Another conjecture that might explain the time consumption of the shooting method is the lengths of the resulting geodesics. The geodesics involving the sweater are those that take longest to compute *and* those that *are* longest. Intuitively, this makes sense, because if the shooting method has to shoot further, the curvature of the manifold should have a greater impact on its performance. For shapes very close to each other, the intermediate space becomes almost flat.

A plot of length vs. time consumption is shown in figure 10. For these geodesics, there is clearly a correspondence. Probably, some more evidence is required to establish a foolproof relation, but the data certainly points to a solution.

8 Final remarks

8.1 Conclusion

We have motivated the study of shape analysis, and presented a representation for shapes as closed or nonclosed curves in $\mathbb{R}^2$, invariant under scaling, rotation and reparametrization. The shape space is an infinite-dimensional manifold. We then undertook the study of geodesics on this manifold, and in so doing, introduced a discretized finite-dimensional manifold, which retains all geodesics from the infinite-dimensional case.

We then presented some results on the existence and uniqueness of geodesics on this manifold, before presenting two methods, one old (the shooting method) and one new (the direct method), for approximating geodesics on the finite-dimensional manifold. Their performance in terms of time, and geodesicity was evaluated, and it is our opinion that the direct method is superior for most reasonably chosen resolutions M, N, but that the shooting method is asymptotically better.

8.2 Future work

Some suggestions for future work include:

- Extend the direct method to the elastic shapes in [14].

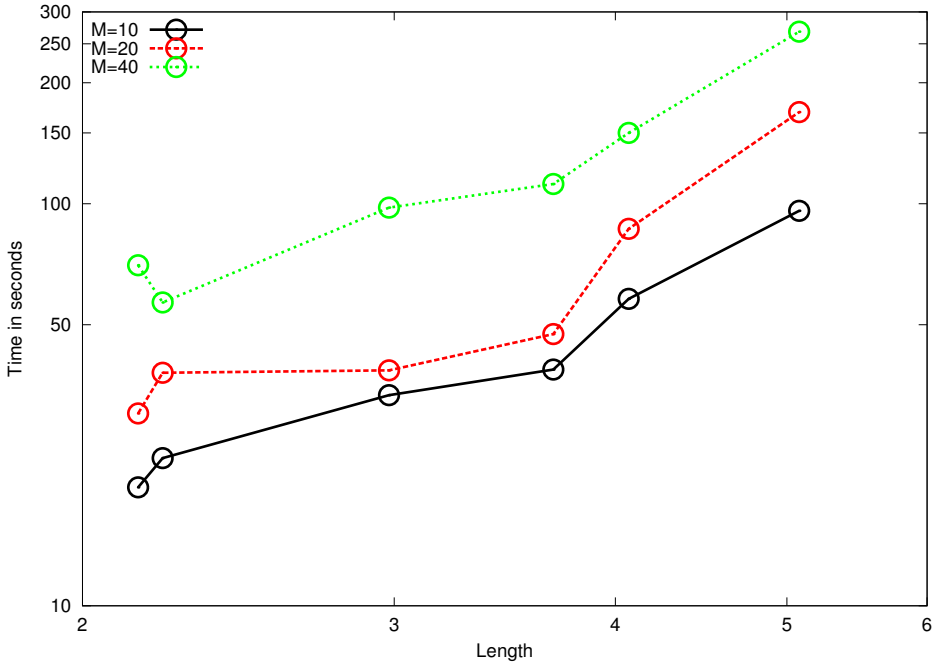


Figure 10: Length of geodesics vs. time to compute with the shooting method, for $M = 10, 20, 40$.

- Compute a lower bound for the injectivity radius $i(x)$ on $\mathcal{C}_c^N$. Is there a positive global lower bound, in spite of $\mathcal{C}_c^N$ not being comapct?
- Are there easily accessible results for the existence and uniqueness of geodesics in $\mathcal{C}_c$?
- Are there results for existence and uniqueness of solutions for the discrete system of equations $Q(X) = 0$?

A Appendix

Definition 13. Let M be a Riemannian manifold with Riemannian metric $\langle \cdot, \cdot \rangle_p$, $p \in M$, and suppose that a Lie group G acts on M with left-multiplication map $L_g : M \to M$ for all $g \in G$. We say that G *acts by isometries* if the pushforward $DL_g|_p : T_pM \to T_{g \cdot p}M$ is an isometry for every $g \in G$ and $p \in M$:

$$\langle DL_g|_p(f), DL_g|_p(g) \rangle_{g \cdot p} = \langle f, g \rangle_p .$$

Theorem 14. *With the setting from definition 13, suppose now that G acts by isometries. Also suppose that G is compact and the action is free, so that M/G is a manifold. Then M/G inherits a natural Riemannian metric from M.*

Proof. Let $[x] \in M/G$. The tangent space at this point is $T_{[x]}M/G \cong T_x^{\perp}[x] \subset T_xM$. The space $T_x^{\perp}[x]$ is called the *horizontal space at* x, and if $\xi \in T_{[x]}M/G$, the *horizontal lift at* x is the unique $\xi_x \in T_x^{\perp}[x]$ satisfying $D\pi|_x(\xi_x) = \xi$, where π is the projection map $M \to M/G$. See for example [1].

Now let $y \in [x]$, $y = g \cdot x$. Clearly, $DL_g|x$ is an isomorphism between $T_x[x]$ and $T_y[x]$, and due to isometry it must also be an isomorphism between $T_x^{\perp}[x]$ and $T_y^{\perp}[x]$. Now, since

$$\pi \circ L_g(x) = \pi(y) = \pi(x)$$

we have

$$D\pi|_y \circ DL_g|_x = D\pi|_x,$$

and specifically,

$$\xi = D\pi|_x(\xi_x) = D\pi|_y \left(DL_g|_x(\xi_x)\right),$$

so that $\xi_y = DL_g|_x(\xi_x)$.

Now let $\xi, \zeta \in T_{[}x]M/G$, and define $\langle \xi, \zeta \rangle_{[x]} = \langle \xi_x, \zeta_x \rangle_x$. This is unambiguous, because if $y \in [x]$, with $y = g \cdot x$ we have, by isometry,

$$\langle \xi_y, \zeta_y \rangle_y = \langle DL_g|_x(\xi_x), DL_g|_x(\zeta_x) \rangle_y = \langle \xi_x, \zeta_x \rangle_x .$$

This gives M/G a Riemannian structure. □

Definition 15. Given a manifold $\mathcal{M}$, a *retraction* at $x \in \mathcal{M}$ is a mapping $\sigma_x : T_xM \to \mathcal{M}$ satisfying $\sigma_x(0) = x$ and $D\sigma_x|_0 = \mathrm{id}_{T_x\mathcal{M}}$, with the canonical identification $T_0T_x\mathcal{M} \cong T_x\mathcal{M}$. We denote by σ the extension of all σ_x to $T\mathcal{M}$, so that $\sigma(p_x) = \sigma_x(p)$. See [19].

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] P. Arias, A. Pini, G. Sanguinetti, P. Sprachmann, P. Cancela, A. Fernández, A. Gómez, and G. Randall. Ultrasound image segmentation with shape priors: Application to automatic cattle rib-eye area estimation. *IEEE Transactions on Image Processing*, 16:1637–1645, 2007.

[3] C. J. Atkin. The Hopf-Rinow theorem is false in infinite dimensions. *Bull. London Math. Soc.*, 7:261–266, 1975.

[4] S. H. Baloch and H. Krim. *Statistics and Analysis of Shapes*, chapter 2D Shape Modeling using Skeletal Graphs in a Morse Theoretic Framework, pages 61–80. Birkhäuser, 2006.

[5] M. Berger. *A Panoramic View of Riemannian Geometry*. Springer-Verlag, 2007.

[6] G. Charpiat, O. Faugeras, and R. Keriven. Approximation of shape metrics and application to shape warping and empirical shape statistics. *Found. Comput. Math.*, 5:1–58, 2005.

[7] C. J. Cotter and D. D. Holm. Singular solutions, momentum maps and computational anatomy. *arXiv:nlin/0605020v1*, 2006.

[8] M. P. do Carmo. *Riemannian Geometry.* Birkhäuser, 1992.

[9] A. Duci, A. Yezzi, S. Soatto, and K. Rocha. Harmonic embeddings for linear shape analysis. *J. Math. Imaging*, 25:341–352, 2006.

[10] H. Hopf and W. Rinow. Über den begriff der völlständigen differentialgeometrischen flächen. *Comment. Math. Helv.*, 3:209–225, 1931.

[11] E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:372–383, 2004.

[12] M. Maeda. On the injective radius of noncompact riemannian manifolds. *Proc. Japan Acad.*, 50:148–151, 1974.

[13] J. Marques and A. Abrantes. Shape alignment - optimal initial point and pose estimation. *Pattern Recognition Letters*, 18:49–53, 1997.

[14] W. Mio, A. Srivastava, and S. Joshi. On shape of plane elastic curves. *International Journal of Computer Vision*, 73:307–324, 2007.

[15] S. B. Myers. Riemanniam manifolds in the large. *Duke Math. J.*, 1:39–49, 1935.

[16] F. Mémoli and G. Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Found. Comput. Math*, 5:313–347, 2005.

[17] A. Prieto-Marquez, P. M. Gignac, and S. Joshi. Neontological evaluation of pelvic skeletal attributes purported to reflect sex in extinct non-avian archosaurs. *Journal of Vertebrate Paleontology*, 27:603–609, 2007.

[18] E. Sharon and D. Mumford. 2d-shape analysis using conformal mapping. *International Journal of Computer Vision*, 70:55–75, 2006.

[19] M. Shub. Some remarks on dynamical systems and numerical analysis. *Dynamical systems and partial differential equations*, pages 69–91, 1984.

[20] J. Zhao, S. Xiong, Y. Bu, and R. S. Goonetilleke. Computerized girth determination for customized footwear manufacture. *Computers & Industrial Engineering*, 54:359–373, 2008.